

周报

2015-06-07

本周总结：

1 数据：

尝试对聊天数据进行入库，才用了单机的Mysql，但是并不理想，导入数据十分缓慢，即使抽取其中一个服务器的数据进行导入，仍然存在过慢（1G数据可能要导入几个小时，数据处理更是要花费很多时间，而事实上聊天数据远远超过1g）因此，对数据算法进行并行并发操作势在必行。

2 算法论文调研：

对课件论文中提到的算法进行了了解，但是跟数据存在问题同样。单纯算法处理并不复杂，但是由于我们数据量较大，因此，普通算法无法满足我们的要求。所以如果着手还是要在并行平台上进行研究。

下周任务：

1 vast 论文review 重读自己的论文，并对review提出的问题进行分析理解

2 寻找合适的并行库并行框架。并尝试将其应用于我们的问题上。